

⑩ 日本国特許庁(JP)

⑪ 特許出願公開

⑫ 公開特許公報(A) 平4-86950

⑬ Int. Cl.⁵

識別記号

庁内整理番号

⑭ 公開 平成4年(1992)3月19日

G 06 F 15/40

5 0 0 U

7056-5L

審査請求 未請求 請求項の数 2 (全5頁)

⑮ 発明の名称 文書検索方法

⑯ 特 願 平2-202704

⑰ 出 願 平2(1990)7月31日

⑱ 発 明 者 寺 村 信 介 東京都大田区中馬込1丁目3番6号 株式会社リコー内

⑲ 出 願 人 株 式 会 社 リ コ ー 東京都大田区中馬込1丁目3番6号

⑳ 代 理 人 弁 理 士 柏 木 明

明 細 書

1. 発明の名称 文書検索方法

2. 特許請求の範囲

1. 利用者に検索キーワードを含む検索条件を入力させ、ファイルに格納された複数の文書中から前記検索条件を満たす所望の文書を検索部により検索するようにした文書検索装置において、検索キーワードを入力する毎に段階的な検索動作を行ない、この検索結果から得られる候補文書に含まれるキーワード情報を利用者に対して表示させるとともに、これらの段階的な検索情報を履歴情報として履歴情報管理部に保存させるようにしたことを特徴とする文書検索方法。

2. 入力された検索キーワード式と履歴情報管理部に保存された過去の履歴情報中のキーワード式とを比較し、過去のキーワード式と完全一致の時には対応する検索結果を前記履歴情報管理部か

ら取出し表示させ、過去のキーワード式と部分一致の時には前記履歴情報管理部に保存された対応する過去の検索結果に対する追加検索を行わせるようにしたことを特徴とする請求項1記載の文書検索方法。

3. 発明の詳細な説明

産業上の利用分野

本発明は、大量の文書中から所望の文書を、複数のキーワードを含む検索条件入力により検索するようにした文書検索方法に関する。

従来の技術

従来、文書検索装置においては種々の検索方式があるが、その一つとして各キーワード間の関連情報を記述したキーワードコネクション表を用いて、関連するキーワードを探し出し、このキーワードから全登録文書について文書確度(利用者の検索要求に近いものほど大きな値を持つ評価値)

を計算し、文書確度の大きい順にソートし検索結果として出力させるようにしたものが、本出願人により例えば特願昭63-8291号「文書検索装置」等により提案されている。このような文書検索装置にあっては、使用者が正確なキーワードを覚えていなくてもキーワードコネクション表を使ったあいまい検索を行なうことにより、適量・適質な検索結果を得ることができる。

このような検索装置にあっては、利用者が適切な文書を検索するための検索条件式として、複数のキーワードをAND、OR又はNOTなどの論理演算子と組合せたキーワード式を用いることができる。しかし、利用者にとって適切に思えるキーワードであってもそのキーワードが必ず登録されているとは限らず、また、本当に適切なキーワード式を利用者が考えることはそれほど容易なことではないので、利用者が望んでいる文書を検索できるまでには無駄が多く、時間もかかって

まれているキーワードが表示される。表示されたキーワードの中に次の検索キーワードとして適切なものがあれば、キーボードなどを用いて入力するか、又は表示部に表示されているキーワードをマウスなどでクリックして選択し入力すればよい。このとき、AND、OR、NOTなどの論理演算子を用いて、最初の検索キーワードとの関係を記述し、キーワード式を構築していく。一方、④は検索結果の文書中に現れるキーワードを出現頻度順に表示するものであり、その中に次の検索キーワードとして適切なものがあれば、上記の場合と同様にキーボード又はマウスにより入力すればよい。

このような段階的検索を行うことで、検索キーワード式の構築を比較的楽に行うことができる。また、このような方法の利点の一つに、段階的に検索を行う際に検索母集団として前回の検索結果を指定することで検索操作を軽くできる。

しまう。

そこで、特願平2-97398号に示されるように、キーワード入力毎に段階的な検索を行い、利用者は検索結果を、以下の

- ① 文書確度の大きい順にソートされた候補文書一覧
- ② 各候補文書の要約
- ③ 各候補の文書中に含まれるキーワード
- ④ 検索結果の文書中に現れたキーワードの統計情報

の何れかの情報形式として見ることができるようにしたものが提案されている。

利用者はこれらの表示内容を参考にし、次の検索キーワードを入力していく。ここに、上記①②は従来の検索システムでも備えられている表示機能であるが、③④は既提案による新規な表示機能である。まず、③は①と関連付けられて表示されるものであり、候補文書名とともにその文書に含

発明が解決しようとする課題

ところが、このような段階的な検索方式によっても、例えば、

- A. ある段階までは適切なキーワード式で検索ができていたが、新しいキーワードを付け加わえて新たに検索を行った結果、その検索キーワード式が適切でなかった（検索キーワード式構築の失敗）。
- B. 過去において用いた検索キーワード式と一部が違うだけで、殆ど同じキーワード式を使って検索操作を行いたい。

というような場合には、全文書に対する検索からやり直さなければならない。一般に、検索は非常にコストの高いオペレーションであり、このような場合には検索効率が著しく悪くなる。特に、Bのような場合、利用者自身が殆ど同じキーワード式で検索を行うことに気が付かないこともあるため、貴重な資源の無駄使いとなる可能性の高いも

のである。

課題を解決するための手段

利用者に検索キーワードを含む検索条件を入力させ、ファイルに格納された複数の文書中から前記検索条件を満たす所望の文書を検索部により検索するようにした文書検索装置において、検索キーワードを入力する毎に段階的な検索動作を行ない、この検索結果から得られる候補文書中に含まれるキーワード情報を利用者に対して表示させるとともに、これらの段階的な検索情報を履歴情報として履歴情報管理部に保存させるようにした。

さらには、入力された検索キーワード式と履歴情報管理部に保存された過去の履歴情報中のキーワード式とを比較し、過去のキーワード式と完全一致の時には対応する検索結果を前記履歴情報管理部から取出し表示させ、過去のキーワード式と部分一致の時には前記履歴情報管理部に保存された対応する過去の検索結果に対する追加検索を行

ード式に対する過去の履歴の利用により自動的に最小のコストで検索ができる。

実施例

本発明の一実施例を図面に基づいて説明する。まず、キーボード1などの入力デバイスや、マウス2などのポインティングデバイスが接続されたキーワード入力部3が設けられている。このキーワード入力部3には入力された検索キーワードを含む検索条件式とともに、キーワードコネクション表(図示せず)やインバーテッドファイル(キーワードから文書へのポインタの集合であるが、特に図示せず)を用いてファイル4中の大量の文書から所望の文書の検索を行なう検索部5が接続されている。また、検索部5により検索された文書情報や、キーワード入力部3からの入力情報等を表示する表示部6が設けられている。さらに、本実施例では、段階的な検索結果及び対応する検索キーワード式等の検索情報を履歴情報として保

わせるようにした。

作用

段階的な検索動作についての検索情報が履歴情報として保存されているので、柔軟な検索操作が可能となり、例えば利用者は段階的に得られる検索結果を見て、最後に入力した検索条件が適切でないと判断した場合には、いつでも前の段階の検索条件に戻って検索をやり直すことができ、効率のよい検索が可能となる。

特に、請求項2記載の発明によれば、検索キーワード式が過去の履歴情報中のキーワード式と完全に一致する場合はもちろん、部分一致の場合であっても過去の検索情報を利用することにより、例えば少しだけ異なるキーワード式を用いて複数の検索を行う場合、最初から検索をやり直す必要がなく、保存された履歴情報中の適切なステージまで戻って検索させることができ、検索操作の負荷を大幅に軽減できる。即ち、入力されたキーワ

存し管理する履歴情報管理部7と、この履歴情報管理部7に保存された段階的な検索結果及び対応する検索キーワード式を取出すための履歴操作部8とが設けられている。また、キーワード入力部3より入力されたキーワード式と、履歴情報管理部7に保存されたキーワード式との比較を行うキーワード比較部9が設けられている。

このような構成において、まず、前提となる既提案の段階的な検索方式を説明する。まず、利用者はキーワード入力部3を通して最初の検索キーワードを入力する。検索部5はこのキーワードとキーワードコネクション表とを用いて検索キーワードリストを作り、これによって文書を検索する。表示部6はこの検索結果を、前述した①～④の何れかの情報形式で選択的に表示する。利用者はこれらの表示内容を参考にし、前述したように次の検索キーワードを入力していく。

具体例で説明する。例えば、利用者がUNIX

に関する文書を検索するため、最初の検索キーワードとして“OS”を入力したとする。この場合の検索結果の上位には、UNIX以外のOS、例えばMS-DOSに関する文書も入っている筈であり、これは③の表示情報から容易に知ることができる。そこで、利用者は次の検索キーワードとして、“AND NOT MS-DOS”と入力する（これは、マウス2を用いて表示部6の表示画面上で“AND”“NOT”“MS-DOS”の順にクリックして選択入力させてもよい）。これにより、“OS AND NOT MS-DOS”というキーワード式を入力したことになる。次の検索対象は、「前回の検索結果の文書」「全文書」の2通りが選べるようにし、利用者が前者で十分だと判断した場合には高速で検索できるものとなる。また、④の表示情報を用いて不要なキーワードの出現頻度が高いときには、これを“NOT”により除外したり、必要なキーワードの出

書を検索する。表示部6は検索結果を上述した①～④の4つの情報の形式で選択的に表示する。同時に、履歴情報管理部7に検索キーワード(式)とこれらの情報とを保存する。利用者は、表示部6で作られた情報を参考にし、次のキーワードを入力し、検索キーワード式を構築する。こうして得られた検索結果と対応するキーワード式はすべて履歴情報管理部7に保存される。履歴操作部8はこの履歴情報を辿ることによって所望の検索段階(ステージ)まで戻り得る機能を持つ。よって、例えば利用者は段階的に得られる検索結果を見て、最後に入力した検索条件式が適切でないと判断した場合には、いつでも前の段階の検索条件式に戻って検索をやり直すことができ、効率のよい検索が可能となる。

ここに、検索対象が前回の検索結果のモードの場合は、必ずキーワード入力部3により入力されたキーワード式が履歴情報管理部7に保存された

現頻度が低いときには容易に指定できる。

このように、既提案によれば、利用者がキーワード式を少しずつ入力する度にインタラクティブに検索過程が表示されるため、適切なキーワード式の構築が容易であり、大量の登録文書からでも利用者が望む文書の絞り込みを効率的に行なうことができる。

しかして、段階的な検索情報を履歴情報として保存するようにした本実施例は「前回の検索結果の文書」を対象とする場合に、特に有効である。まず、利用者はキーワード入力部3によりキーワードを入力する。最初のキーワードの入力としては、キーボード1による直接入力と、表示部6にキーワード一覧を出力させてこの一覧の中からマウス2によって選ぶ入力との2通りがある。何れにしても、検索部5は入力されたキーワードとキーワードコネクション表を用いて前述したように検索キーワードリストを作り、これによって文

履歴情報中のキーワード式とマッチングするかがキーワード比較部9で比較される。もし、過去において全く同じキーワード式(完全一致)による検索が行われていたら、対応する検索結果情報を履歴情報管理部7から取出して表示部6に表示させる。一方、完全にはマッチしない場合であっても、入力したキーワードの部分式になっているもの(部分一致)の内、最も長いものを探してその結果に対して式の差分の追加検索を行わせる。このように、入力されたキーワード式に対して過去の履歴を利用して自動的に最小コストで検索可能となる。例えば、入力キーワード式がa & b & c & dであり、過去においてa & b及びa & b & cなるキーワード式による検索が行われていれば、後者の検索結果に対してdなる条件を加えて検索を行えばよいことになる。

発明の効果

本発明は、上述したように段階的な検索動作に

についての検索情報を履歴情報として履歴情報部に保存させるようにしたので、柔軟な検索操作が可能となり、例えば利用者は段階的に得られる検索結果を見て、最後に入力した検索条件が適切でないと感じた場合には、いつでも前の段階の検索条件に戻って検索をやり直すことができ、効率のよい検索が可能となり、特に、請求項2記載の発明によれば、検索キーワード式が過去の履歴情報中のキーワード式と完全に一致する場合はもちろん、部分一致の場合であっても過去の検索情報を利用するようにしたので、例えば少しだけ異なるキーワード式を用いて複数の検索を行う場合、最初から検索をやり直す必要がなく、保存された履歴情報中の適切なステージまで戻って検索させることができ、検索操作の負荷を大幅に軽減でき、入力されたキーワード式に対する過去の履歴の利用により自動的に最小のコストで検索ができるものである。

4. 図面の簡単な説明

図面は本発明の一実施例を示すブロック図である。

4…ファイル、5…検索部、6…表示部、7…履歴情報管理部、9…比較部

出願人 株式会社 リコー
代理人 柏 木

